# Contrast Feature Dependency Pattern Mining for Controlled Experiments with Application to Driving Behavior

Qingzhe Li, Liang Zhao, Yi-Ching Lee, Yanfang Ye, Jessica Lin, Lingfei Wu

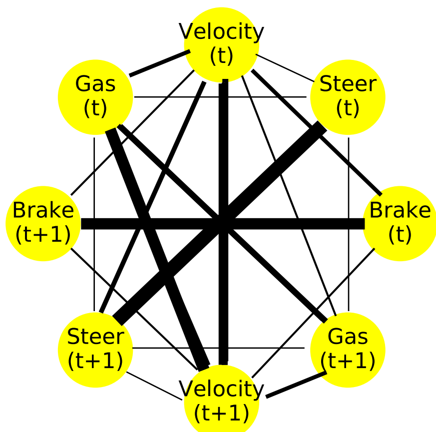
**Presented by Qingzhe Li**

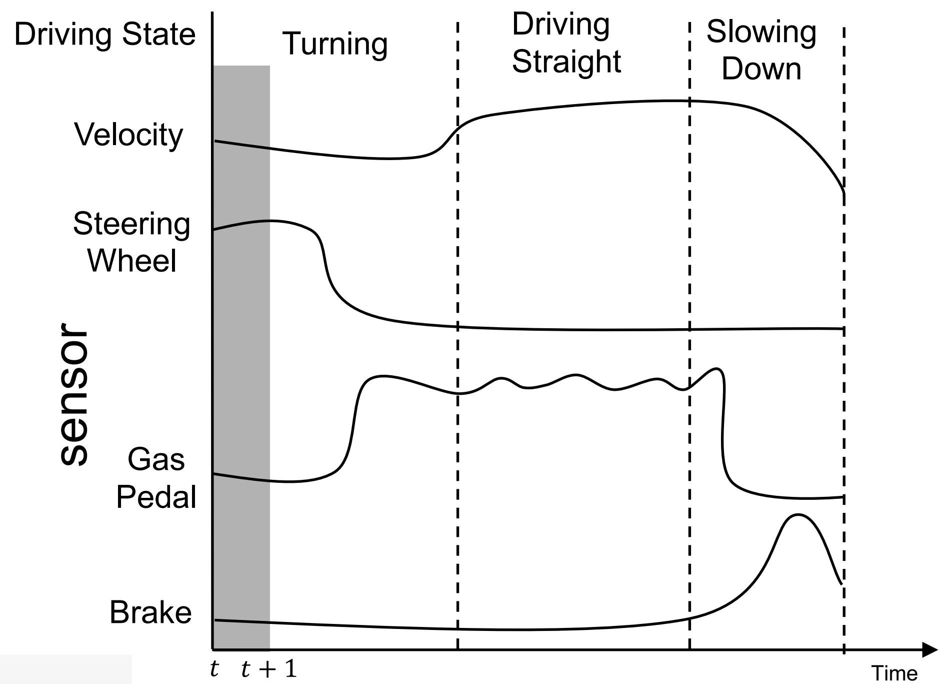
GEORGE MASON UNIVERSITY

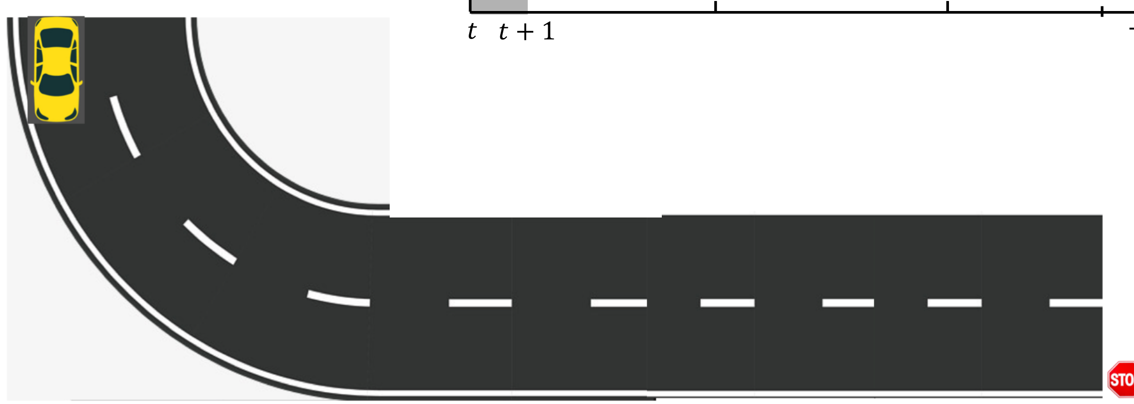
# Introduction

- Charactering the latent state in multivariate time series

Dependency Network of

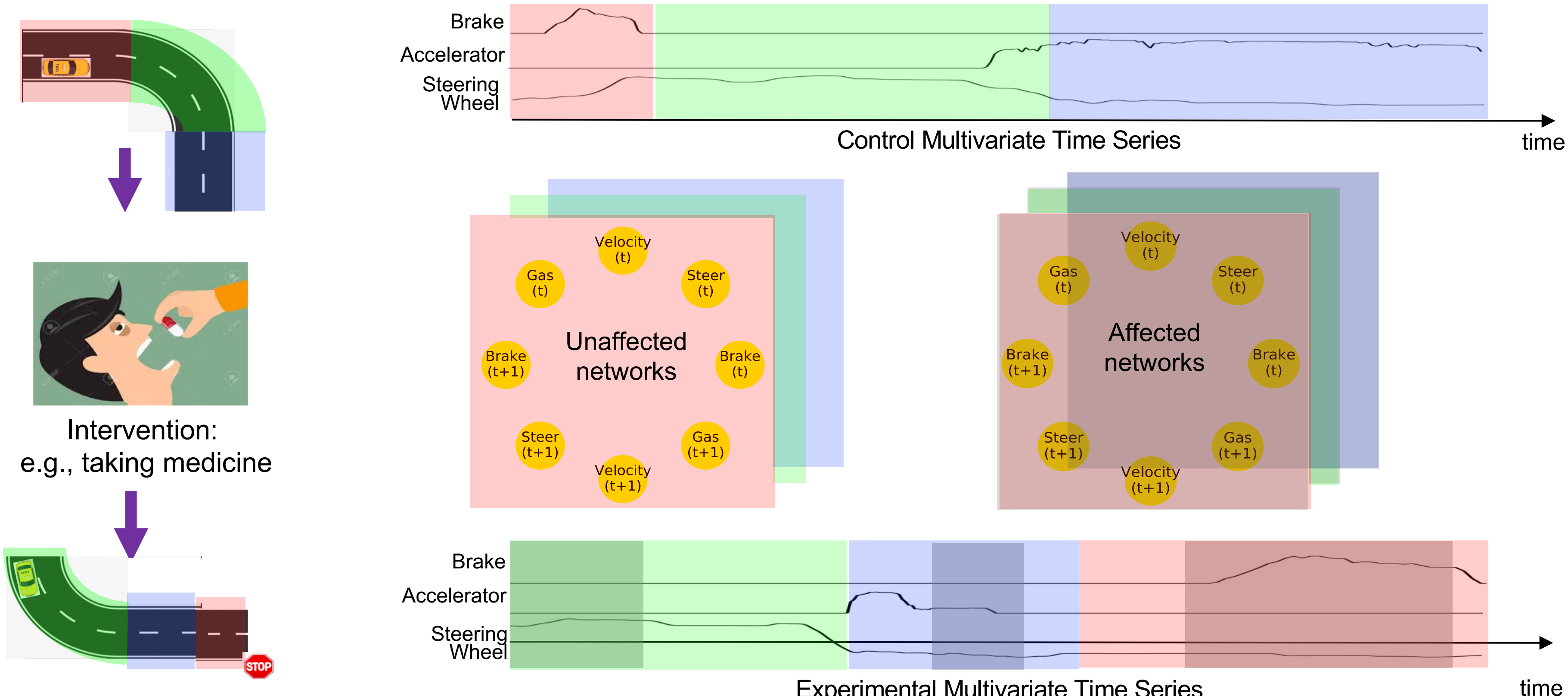


# Contrast Pattern in Controlled Experiments

- **Goal: determine which the driving behaviors are affected or not**




- **Problem formulation**

Inputs: 1. Control Multivariate Time Series 2. Experimental Multivariate Time Series

Outputs: 1. Latent state assignments 2. Contrast pattern detection 3. Contrast pattern characterization

# Challenges

1. Integrally modeling the coupled outputs:

Outputs: 1. Latent state assignments 2. Contrast pattern detection 3. Contrast pattern characterization

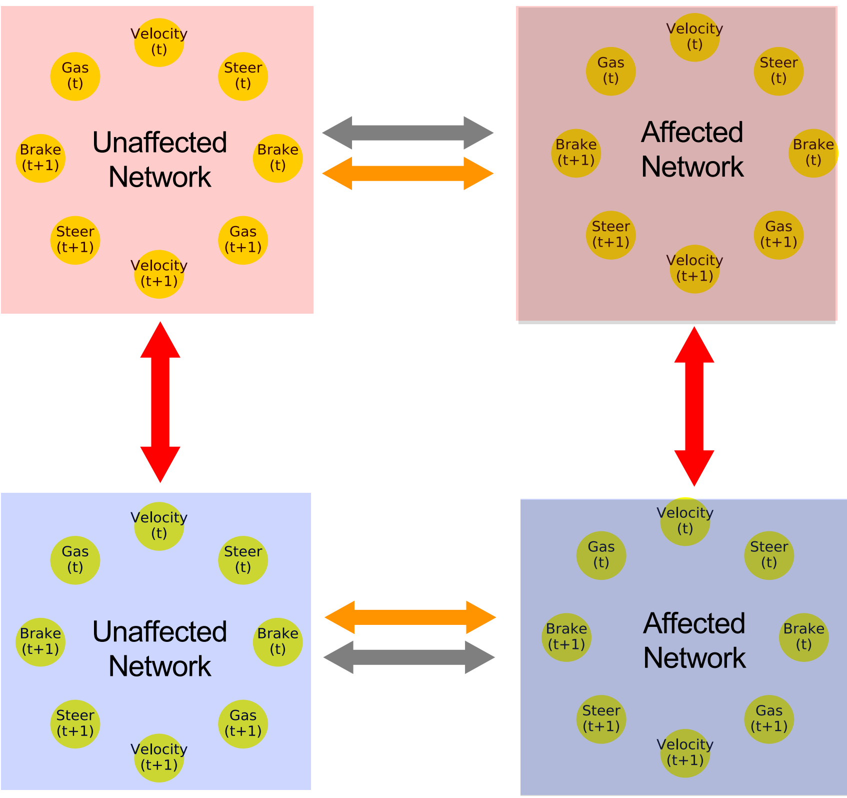
Coupled

2. Differentiating the dependency networks:

Similarities caused by same latent state

Differences caused by intervention

Differences caused by latent state



# Resolving Challenge 1

**Input:**

1. Control MTS: $X$
2. Experimental MTS: $\hat{X}$

Coupled

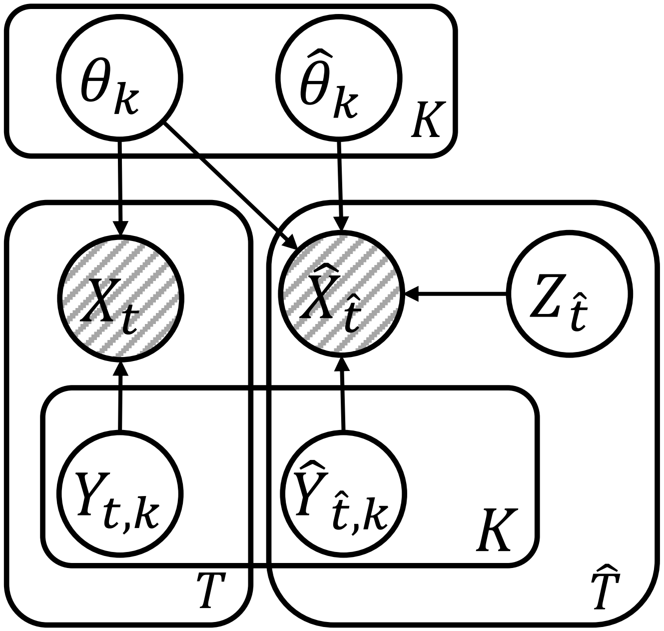
**Output:**

1. Latent state assignments: $Y, \hat{Y}$
2. Contrast pattern detection: $Z$
3. Contrast latent state characterization: $\theta = \{\theta_k\}_{k=1}^{K}, \hat{\theta} = \{\hat{\theta}_k\}_{k=1}^{K}$

- Proposing an integrated generative model for contrast pattern mining problem

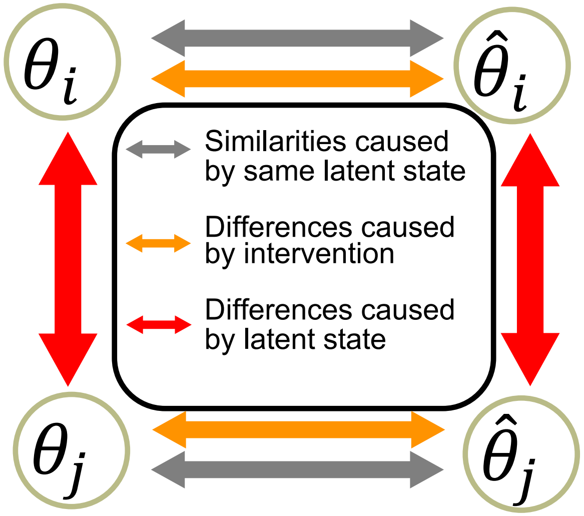
- Joint likelihood

$$p(X, \hat{X}|Y, Z, \theta, \hat{\theta}) = p(X|Y, \theta) \cdot p(\hat{X}|\hat{Y}, Z, \theta, \hat{\theta})$$



# Resolving Challenge 2

- **Intuitions / Motivations**
  - The latent states are decided by environments
  - The contrast patterns are decided by intervention
  - The intervention is unlikely to change the latent states, i.e., $Similarity(\theta_i, \hat{\theta}_i) > Similarity(\theta_i, \theta_j)$




- **Technical challenge**
  - Flawed definition of similarity directly between two inverse covariance matrices, for example:
    - Non-interpretable single element in $\theta_i$
    - Different scales between $\theta_i$ and $\theta_j$
- **Proposing Partial Correlation Based Regularization**

$$\mathcal{R}_{\mathcal{C}}(\theta, \hat{\theta}) = \lambda \cdot \sum_{k}^{K} \|\rho_k - \hat{\rho}_k\|_F^2$$

where $\rho_k$ and $\hat{\rho}_k$ are the partial correlation matrices computed from $\theta_i$ and $\hat{\theta}_k$

# Overall Objective Function

Objective:

Loss function: negative log likelihood

Partial Correlation Based Regularization

$$\arg\min_{\theta,\hat{\theta},Y,\hat{Y},Z} \quad \mathcal{L}(Y,\hat{Y},Z,\theta,\hat{\theta}) + \mathcal{R}_{\mathcal{C}}(\theta,\hat{\theta}) + \mathcal{R}_{\mathcal{T}}(Y,\hat{Y},Z)$$

Smoothening the latent state assignments between the neighbor time indices

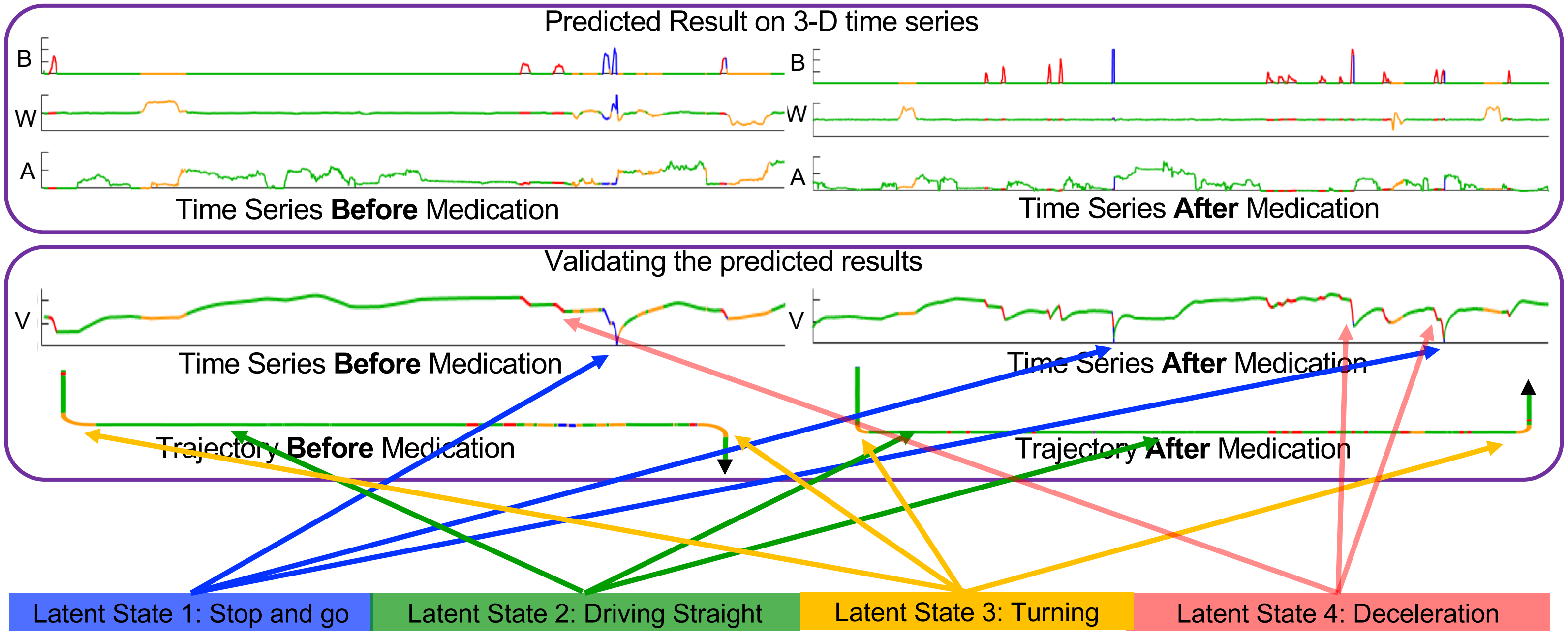
where $\mathcal{R}_{\mathcal{T}}(Y,\hat{Y},Z) = \sum_{t=2}^{T} \gamma \mathbb{1}(Y_t \neq Y_{t-1}) + \sum_{\hat{t}=2}^{\hat{T}} \beta \mathbb{1}(Z_{\hat{t}} \neq Z_{\hat{t}-1}) + \gamma \mathbb{1}(\hat{Y}_{\hat{t}} \neq \hat{Y}_{\hat{t}-1})$

Optimization Algorithm:

- Repeat
  - Expectation-step: fix continuous variables $(\theta, \hat{\theta})$ optimize discrete variables $(Y, Z)$
    - By formulating a dynamic programming problem
  - Maximization-step: fix $Y, Z$ optimize $\theta, \hat{\theta}$
    - By developing an ADMM based algorithm
- Until Stationarity

# Experimental Result on Real-world Datasets

- Predicted Latent State Assignments



# Experimental Result on Real-world Datasets (cont.)

- Visualizing the learned contrast patterns

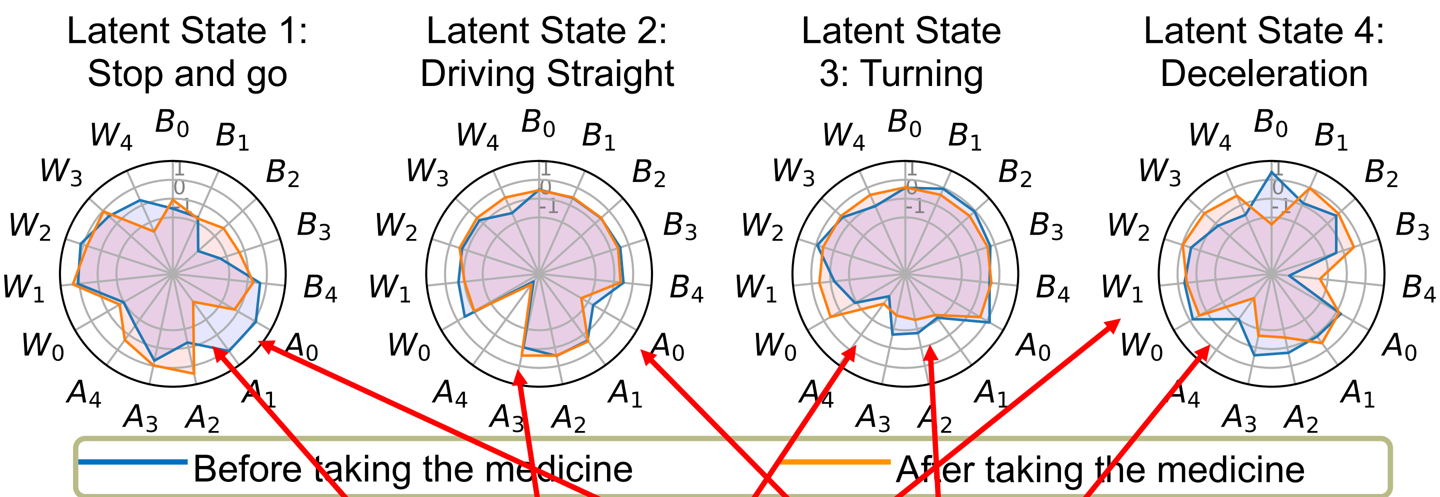



Figure: Closeness centrality scores of the learned dependency networks. The higher score of the node, the more important of the node in the network

The non-overlapped areas indicate the driving behaviors have been changed by the medicine.

For the same latent state, the centrality scores are similar of all nodes

# QUESTIONS?

# Experimental Result on Synthetic Datasets

TABLE II: The (macro) $F_1$ scores of predicted $Y, \hat{Y}$, and $Z$ assignments

| Method | Dataset 1 $Y$ | Dataset 1 $\hat{Y}$ | Dataset 1 $Z$ | Dataset 2 $Y$ | Dataset 2 $\hat{Y}$ | Dataset 2 $Z$ | Dataset 3 $Y$ | Dataset 3 $\hat{Y}$ | Dataset 3 $Z$ |
|---|---|---|---|---|---|---|---|---|---|
| K-means+1SVM | 0.50 | 0.51 | 0.58 | 0.33 | 0.34 | 0.61 | 0.28 | 0.27 | 0.60 |
| K-means+EE | 0.50 | 0.51 | 0.23 | 0.33 | 0.34 | 0.25 | 0.28 | 0.27 | 0.25 |
| K-means+IF | 0.50 | 0.51 | 0.23 | 0.33 | 0.34 | 0.26 | 0.28 | 0.27 | 0.26 |
| K-means+LOF | 0.50 | 0.51 | 0.15 | 0.33 | 0.34 | 0.18 | 0.28 | 0.27 | 0.21 |
| K-shape+1SVM | 0.51 | 0.51 | 0.54 | 0.34 | 0.33 | 0.56 | 0.26 | 0.24 | 0.55 |
| K-shape+EE | 0.51 | 0.51 | 0.23 | 0.34 | 0.33 | 0.25 | 0.26 | 0.24 | 0.25 |
| K-shape+IF | 0.51 | 0.51 | 0.24 | 0.34 | 0.33 | 0.26 | 0.26 | 0.24 | 0.25 |
| K-shape+LOF | 0.51 | 0.51 | 0.14 | 0.34 | 0.33 | 0.19 | 0.26 | 0.24 | 0.21 |
| TICC+1SVM | **0.99** | 0.72 | 0.47 | 0.29 | 0.24 | 0.48 | 0.25 | 0.23 | 0.51 |
| TICC+EE | 0.99 | 0.72 | 0.35 | 0.29 | 0.24 | 0.25 | 0.25 | 0.23 | 0.25 |
| TICC+IF | 0.99 | 0.72 | 0.29 | 0.29 | 0.24 | 0.27 | 0.25 | 0.23 | 0.25 |
| TICC+LOF | 0.99 | 0.72 | 0.30 | 0.29 | 0.24 | 0.20 | 0.25 | 0.23 | 0.25 |
| GMM+1SVM | 0.95 | 0.87 | 0.49 | 0.85 | **0.80** | 0.50 | 0.83 | **0.78** | 0.52 |
| GMM+EE | 0.95 | 0.87 | 0.22 | 0.85 | 0.80 | 0.22 | 0.83 | 0.78 | 0.24 |
| GMM+IF | 0.95 | 0.87 | 0.23 | 0.85 | 0.80 | 0.24 | 0.83 | 0.78 | 0.25 |
| GMM+LOF | 0.95 | 0.87 | 0.16 | 0.85 | 0.80 | 0.18 | 0.83 | 0.78 | 0.21 |
| Baseline ($\lambda = 0$) | 0.94 | **0.92** | **0.89** | **0.86** | 0.63 | **0.88** | **0.83** | 0.59 | **0.76** |
| **CPM-P (ours)** | **0.99** | **0.99** | **0.98** | **0.99** | **0.75** | **0.89** | **0.99** | **0.99** | **0.98** |

Only use loss function without proposed regularization outperform comparison

Our method successfully identified the latent states and the contrast patterns with high $F_1$ score

Methods without considering the coupling between the latent state and contrast patterns does NOT work